Pixel-Level Face Image Quality Assessment for Explainable Face Recognition

Philipp Terhörst, Marco Huber, Naser Damer, *Member, IEEE*, Florian Kirchbuchner, *Member, IEEE*, Kiran Raja, *Senior Member, IEEE*, and Arjan Kuijper, *Member, IEEE*

***Abstract*—In this work, we introduce the concept of pixel-level face image quality that determines the utility of single pixels in a face image for recognition. We propose a training-free approach to assess the pixel-level qualities of a face image given an arbitrary face recognition network. To achieve this, a model-specific quality value of the input image is estimated and used to build a sample-specific quality regression model. Based on this model, quality-based gradients are back-propagated and converted into pixel-level quality estimates. In the experiments, we qualitatively and quantitatively investigated the meaningfulness of our proposed pixel-level qualities based on real and artificial disturbances and by comparing the explanation maps on faces incompliant with the ICAO standards. In all scenarios, the results demonstrate that the proposed solution produces meaningful pixel-level qualities enhancing the interpretability of the face image and its quality. The code is publicly available.**

***Index Terms*—Biometrics, quality assessment, explainable face recognition, interpretable face recognition.**

I. Introduction

FACE recognition (FR) systems are spreading worldwide and have a growing effect on our daily lifes [40]. Since these systems are increasingly involved in critical decision-making processes, such as in forensics and law enforcement, there is a growing need in making the FR process explainable to humans [31]. Especially in unconstrained environments, FR systems have to deal with large variabilities, such as image acquisition conditions (illumination, background) and factors of the face (pose, occlusions, expressions), that might result in defective matching decisions [20], [21]. The impact of these variabilities on the FR performance is measured in terms of face image quality (FIQ). Consequently, the performance of FR systems is strongly dependent on the quality of their samples. The FIQ of a sample is defined as its utility for recognition [5], [13], [17], [29]. The automatic prediction of FIQ (prior to matching) is one of the key factors during the enrolment and is essential to achieve robust and accurate FR performances [31].

Previous research on FIQ focused mainly on the development of accurate quality assessment methods [5], [7], [17], [28], [39]. Although these methods possess similar bias problems than for FR systems [38], no works aimed at making the output of FIQ assessment (FIQA) methods explainable to humans [31], and thus provide an interpretable reason for a face image being of low or high quality. On the other hand, previous works on explainable FR focused solely on making the matching decision explainable to humans, neglecting that explainability is also needed during the enrolment of subjects, where the quality compliance of the image is typically checked.

In this work, we propose a training-free approach to compute pixel-level quality (PLQ) explanation maps that determines the utility of single pixels for recognition, similar to the definition of FIQ. The PLQ-maps aim at making the enrolment process explainable for humans. Their construction consists of three steps as shown in Figure 1. First, a model-specific quality value for an input image is estimated. Second, this quality value and the FR model are used to build a sample-specific quality regression model without the need for training. Third, this quality model, optimized for the input image, is used to back-propagate quality-based gradients and convert these into PLQ estimates.

In the experiments, the effectiveness of the proposed PLQ-maps are evaluated quantitatively and qualitatively in three scenarios. This is done by demonstrating that areas of low pixel-quality result in lower FIQ values and vice versa. First, it is shown that inpainting low pixel-quality areas in the face (such as occlusions) localised by our method increases the FIQ. Second, it is demonstrated that placing random disturbances on the face results in easily-detectable areas of low pixel-quality. Third, the PLQ-maps are analysed based on face images incompliant to various International Civil Aviation Organization (ICAO) specifications [20]. In all three scenarios, the results demonstrate that the proposed solution produces meaningful PLQ values.

Manuscript received 5 September 2022; revised 5 January 2023; accepted 27 March 2023. Date of publication 30 March 2023; date of current version 16 May 2023. This work was supported by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE. This article was recommended for publication by Associate Editor N. Evans upon evaluation of the reviewers' comments. *(Corresponding author: Philipp Terhörst.)*

Philipp Terhörst is with the the Department of Computer Science, Norwegian University of Science and Technology, 7034 Trondheim, Norway, also with the Department of Smart Living & Biometric Technologies, Fraunhofer Institute for Computer Graphics Research IGD, 64283 Darmstadt, Germany, and also with the Department of Computer Science, Paderborn University, 33098 Paderborn, Germany (e-mail: Philipp.Terhoerst@uni-paderborn.de).

Marco Huber, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper are with the Department for Smart Living & Biometric Technologies, Fraunhofer Institute for Computer Graphics Research IGD, 64283 Darmstadt, Germany, and also with the Department of Computer Science, Technical University of Darmstadt, 64289 Darmstadt, Germany (e-mail: Marco.Huber@igd.fraunhofer.de; Naser.Damer@igd.fraunhofer.de; Florian.Kirchbuchner@igd.fraunhofer.de; Arjan.Kuijper@igd.fraunhofer.de).

Kiran Raja is with the Department of Computer Science, Norwegian University of Science and Technology, 7034 Trondheim, Norway (e-mail: Kiran.Raja@ntnu.no).

Digital Object Identifier 10.1109/TBIOM.2023.3263186

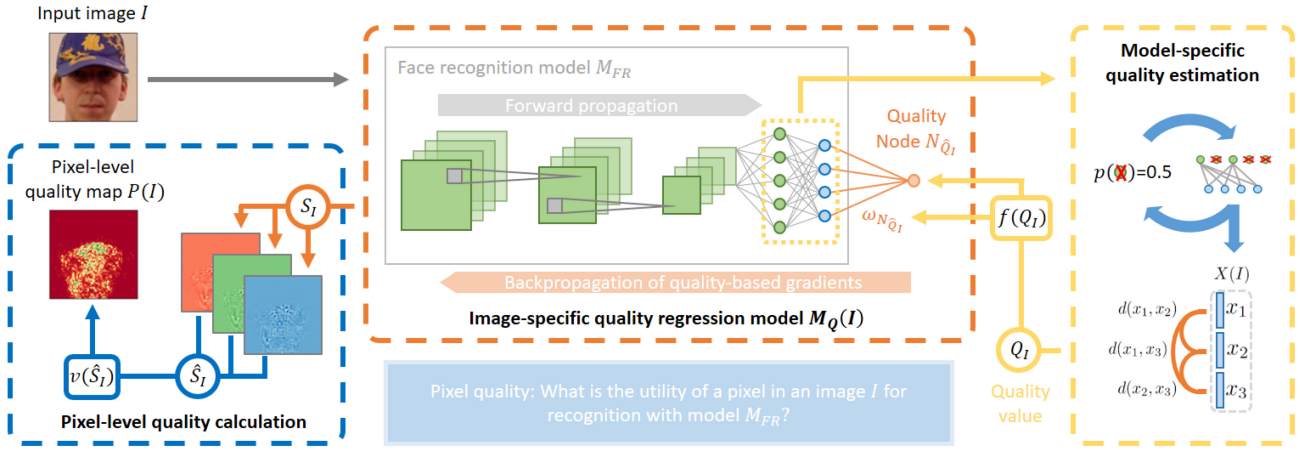


Fig. 1. **Overview of the proposed pixel-level quality estimation approach** in three steps. First, the input image I is passed through a face recognition model $\mathcal{M}_{FR}$ with repetitive forward-passes through the last layer to obtain a model-specific quality estimate Q_I. This quality value is modified and used in the second step to build an image-specific quality regression model $\mathcal{M}_Q(I)$ by extending $\mathcal{M}_{FR}$ with a quality node $N_{\hat{Q}_I}$ that is connected through the weights $\omega_{N_{\hat{Q}_I}}$. In the third step, the constructed regression model $\mathcal{M}_Q(I)$ is used to backpropagate the quality-based gradients and to transform the resulting saliency maps to a pixel-level quality map $P(I)$.

The proposed PLQ-maps have several advantages. First, they can be used to deepen the understanding of how FIQA and FR models work since these maps describes the importance of pixels for the model-specific FIQ and therefore also for the FR model. Second, they can be used to enhance FR performance by inpainting or merging low-quality areas of the face to create images of higher utility. Last, they can explain why an image cannot be used as a reference image during the acquisition/enrolment process and in which area of the face the subject have to do changes to increase the quality. Consequently, PLQ maps provide guidance on the reasons behind low quality images, and thus can provide interpretable instructions to improve the FIQ.

To summarize, the proposed PLQA approach (a) can be applied on arbitrary FR networks, (b) does not require training, and (c) provides a pixel-level utility description of an input face explaining how well pixels in face image are suited for recognition before it is used for matching. The code for this work is publicly available.[1]

II. Related Work

A. Explainable Face Recognition

Explainable FR is a relatively new field of research that aims at making the face recognition pipeline, and its consequences, explainable for humans. In 2019, Yin et al. [45] proposed a spatial activation diversity loss. The loss penalizes correlations among filter weights and they showed that their filter distribution is more spread to different spatial areas. This lead to learned face representations of higher structure and therefore higher interpretability since each dimension of the representation represents a face structure or a face part. In [46], Zee et al. trained a classification network on faces and used class activation maps to find the most distinguishable regions. With this information, the authors showed that the human FR performance is increased. In 2020, Williford et al. [42] proposed new approaches for explainable FR. Based on triplets consisting of a probe, a mate, and a non-mate image, the algorithms generate saliency maps that highlight the maximum similarity between the probe and the mate and the minimum between the probe and the non-mate. This provides explanations on why the matcher comes to a certain decision.

So far, works on explainable FR have focused on making the matching decision explainable. Contrarily, in this work, we propose a method to make the utility of an image for recognition explainable before any matchings.

B. Face Image Quality Assessment

Several standards have been proposed to ensure face image quality by constraining the capture requirements, such as ISO/IEC 19794-5 [21] and ICAO 9303 [20]. These standards divide quality into *image-based* qualities (such as illumination, occlusion) and *subject-based* quality measures (such as pose, expression, accessories). This influenced the first generation of FIQA approaches that are built on human perceptive image quality factors [1], [2], [10], [12], [13], [18], [29], [33], [41]. However, due to the achieved performance, the research focus shifted to learning-based approaches.

The second generation of FIQA approaches [3], [5], [7], [17], [23], [43] consists of supervised learning algorithms based on human or artificially constructed quality labels. These quality labels were either based on human judgement or derived from comparison score distributions. The utilized algorithms include rank-based learning [7], the use of SVM-based approaches [5], and training deep networks with artificial quality labels [17], [28], [44]. However, humans may not know the best characteristics for face recognition systems and artificially labelled quality values, derived from comparison scores, rely on error-prone labelling mechanisms and require large-scale training.

The third generation of FIQA approaches completely avoids the use of quality labels. In 2020, Terhörst et al. [39] proposed stochastic embedding robustness for FIQA (SER-FIQ). This

[1] https://github.com/pterhoer/ExplainableFaceImageQuality

concept measures the robustness of a face representation against dropout variations and uses this measure to determine the quality of a face. It avoids the need for training and takes into account the decision patterns of the deployed face recognition model. In 2021, Meng et al. [26] proposed a class of loss functions that include magnitude-aware angular margins, encoding the quality into the face representation. Training with this loss results in an FR model that produces embeddings whose magnitudes can measure the FIQ of their faces.

So far, research on FIQ focuses only on the development of FIQA methods. Although that it was shown that FIQA possesses similar bias problems than for FR [38], no works aimed at making the output of FIQA explainable to humans. While there are various approaches to visualize classification decisions of deep learning models [27], [34], [37], to the best of our knowledge, this is the first work on explaining the utility of face representations.

III. Methodology

The proposed pixel-level quality estimation method consists of three steps. First, for a given face image I, a model-specific quality estimate is computed stating its utility for the face recognition network. Second, the quality value and the recognition network are used to build a quality regression model without the need for training. Third, this model is used to back-propagate quality-based gradients and convert these into pixel-level face image quality estimations. An overview of the proposed concept is shown in Figure 1.

A. Model-Specific Quality Estimation

To compute the model-specific FIQ value Q_I, our method builds on the work of Terhörst et al. [39]. This choice is based on its training-free applicability to arbitrary FR networks and since it determines how well a specific model $\mathcal{M}_{FR}$ can use I for recognition. Given a face image I, this image is propagated through the network and the forward passes to the last (embedding) layer are repeated $m = 100$ times as motivated in [39]. During each of these stochastic forward passes, a different dropout pattern (with $p_d = 0.5$) is applied resulting in a set of m different stochastic embeddings X_I. The FIQ of the image I is given by

$$Q_I = Q(X_I) = 2\sigma\left(-\frac{2}{m^2}\sum_{i<j} d(x_i, x_j)\right), \tag{1}$$

where $\sigma(\cdot)$ is the sigmoid function and states the Euclidean distance between two stochastic embedding $x_i, x_j \in X_I$. Q_I defines the quality of an image over the robustness of its embeddings. If there are high variations in the stochastic embeddings, the robustness of the representation is low and thus the quality. Since the quality score is model-dependent and often in a narrow range, we additionally adjust the score to the range of [0, 1] using

$$\hat{Q}_I = f(Q_I) = \sigma(\alpha(Q_I - r)). \tag{2}$$

Choosing r near the mean of the quality distribution of a development set ensures a new mean quality around 0.5 after applying Eq. (2). Parameter α is chosen to stretch the values to a range of [0, 1]. Please note that this quality scaling is optional and only aims at making the results more easily comparable.

B. Building a Quality Regression Model

Based on the face image quality score $\hat{Q}_I$ for image I and the face recognition model $\mathcal{M}_{FR}$, we now build an image-specific quality regression model $\mathcal{M}_Q$ in a training-free fashion based on this single input image. This is performed by extending the face recognition model $\mathcal{M}_{FR}$ with a one-dimensional quality node $N_{\hat{Q}_I}$. The node is fully connected to the (last) embedding layer of $\mathcal{M}_{FR}$. The weights of these connections are given by

$$w_{N_{\hat{Q}_I}} = \frac{\hat{Q}_I}{||e_I||_1}, \tag{3}$$

where $\mathcal{M}_{FR}(I) = e_I$ is the face embedding of $\mathcal{M}_{FR}$ for the single image I. This assumes a linear layer activation with a bias term of $b = 0$ and ensures that all features of e_I are equally important for the quality estimation. Moreover, the construction of $\mathcal{M}_Q$ ensures that given the single image I, the output of the model is $\hat{Q}_I$.

C. Pixel-Level Quality Calculation

The constructed quality regression model $\mathcal{M}_Q$ for image I is, similarly to $\mathcal{M}_{FR}$, pairwise differentiable. Therefore, we can compute a gradient-based saliency map

$$S(I) = \frac{\delta \mathcal{M}_Q(I)}{\delta I}, \tag{4}$$

similar to [4], [35], [36]. In contrast to these, the saliency map in our work is not dependent on a certain class but rather on the continuous quality value. The saliency map $S(I)$ consists of the gradients for each pixel of I. The magnitudes of these gradients indicate the relative effect on each pixel on the FIQ value.

Considering $S(I)$, only the magnitudes of the gradients are crucial for the quality assessment task while their directions are context-dependent [36]. Consequently, the three color channels of $S(I)$ are merged considering only the absolute values of the gradients. This is done by

$$\hat{S}(I) = \frac{1}{3}\sum_{c=1}^{3} |g_{i,j,c}|, \tag{5}$$

where $g_{i,j,c}$ represents the gradient for pixel (i, j) of color channel c. $\hat{S}(I)$ can already be interpreted as pixel-level qualities. However, since the pixel-level qualities aim at visually explaining the utility of an image for recognition in a human-understandable manner and the ranges of $\hat{S}(I)$ are, depending on $\mathcal{M}_{FR}$, in a narrow range, a visualization function v

$$v\left(\hat{S}\right) = 1 - \frac{1}{1 + \left(10^{\gamma} \times \hat{g}^2_{i,j,c}\right)}, \tag{6}$$

is used to project $\hat{S}$ to a more intuitive range of [0, 1]. The visualization parameter γ is used to stretch the quality values to the desired range. Applying $v(\cdot)$ on $\hat{S}(I)$ results in the pixel-level quality map $P(I) = v(\hat{S}(I))$. $P(I)$ is the representation of the pixel-level qualities $p_{i,j} \in [0, 1]$. A higher pixel quality

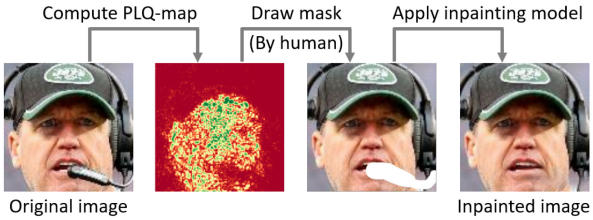


Fig. 2. **Workflow of the Pre- and Inpainted Dataset creation** - Using the proposed methodology, the PLQ-map is computed. Then, both images are given to a human observer to mask the quality-decreasing factor according to the interpretation of the observer. Then, this area is inpainted [24] to create a similar face image without the potential quality-decreasing factor. The difference between the FIQ of the original and the inpainted image allows to state the interpretability of the PLQ-map for the observer.

indicates a higher contribution for the recognition utility of the face image and vice versa.

Please note that, in contrast to the typical procedure when dealing with gradient backpropagation, we do not scale the gradients per image, e.g., with MinMax scaling. Scaling the gradients would highlight the differences in low- and high-quality regions of a single image but also result in the loss of the PLQ comparability between different images. For instance, such scaling will always result in an area of high quality even if the image is not suitable for recognition.

IV. Experimental Setup

A. Databases

1) The Pre- and Inpainted Dataset: Was created by manually selecting images from VGGFace2 [6] and Adience [11] since these contain images of large variances. The decision criteria for the selection was that the images must contain occlusions or similar quality-decreasing factors according to human judgement. For each image, the PLQ-map is computed based on the proposed approach and both images, the original and the PLQ-map, are given to a human observer with the task of determining and masking the quality-decreasing factor. The image and the corresponding mask are given to an inpainting model [24]. This resulted in pairs of similar face images with and without quality-decreasing factors. When the FIQ of the inpainted image is higher than the original FIQ, then we can conclude that the PLQ-maps can be successfully interpreted to detect low-quality areas. The workflow of the dataset creation is shown in Figure 2. The created dataset consists of 100 pairs of face images with their inpainted counterparts.

2) The Random Mask Dataset: Is based on the ColorFeret [30] database due to the high image-quality of its images which corresponds to scenarios such as identity document and border checks. Each image of the random mask dataset was created by placing a black square on the inner image of a frontal face. For each ColorFeret image, 5 black squares of size $s \times s$ ($s = 10, 20, 30, 40, 50$) pixels were placed randomly on the image resulting in 5 images for the random mask dataset. To avoid that non-facial areas are masked, the squares are only placed in the inner 90% of the face images. This results in a total of 6610 masked face images to demonstrate that the proposed methodology can detect these disturbances as low-quality regions.

3) The Inhouse ICAO Incompliance Dataset: Was collected by us to analyse the effect of pixel-level face image quality on face images that violate various International Civil Aviation Organization (ICAO) specifications [20]. It consists of a reference image of one subject that complies with these specifications as well as 33 face images of the same subject with different violations of these specifications. The images were taken with fixed capturing conditions to allow a clear investigation of the effect of pixel-level face image qualities on ICAO incompliances.

B. Face Recognition Models and Parameters

We analysed the image- and pixel-level face quality based on two widely-used FR models using FaceNet[2] [32] and ArcFace[3] [8] losses (both MIT License) based on ResNet-100. For the sake of simplicity, we refer to these models as FaceNet and ArcFace. Both models were trained on the MS1M database [16]. Given a face image, the image is aligned, scaled, and cropped before being passed to one of the models. This preprocessing is done as described in [15] for ArcFace, and as described in [22] for FaceNet.

Since the quality estimations are model-specific for FR systems, the parameters for the quality scaling and the quality visualization are as well. The quality values are adjusted to a wider range of $[0, 1]$ on the quality values of the Adience benchmark [11] and resulted in parameters $\alpha_{AF} = 130$, $r_{AF} = 0.88$ for ArcFace and $\alpha_{FN} = 450$, $r_{FN} = 0.93$ for FaceNet. For visualizing the pixel-level qualities, we choose $\gamma_{AF} = 7.5$ for ArcFace and $\gamma_{FN} = 5.5$ for FaceNet. Please note that the choice of γ is subjective and depending on the colormap used for visualizing the quality values.[4] In general, the choice of these parameters (α, r, γ) determine the scaling of the qualities and thus, aim to make the results more easily understandable. Since the scaling is done with a strictly increasing function, the order of the qualities, and thus the FIQA task in general, is not affected.

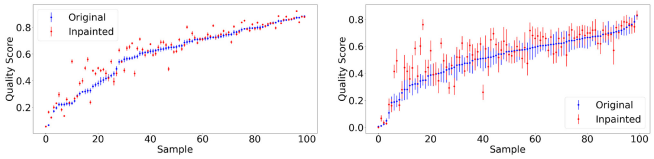
C. Investigations

The proposed approach is analysed in three steps. First, the human interpretability of the PLQ-maps is investigated by giving a human observer low-quality face images and the corresponding PLQ-maps to localise the quality-decreasing factor. Second, random masks are placed on high-quality faces to show that the proposed methodology identifies these as low-quality areas. Both evaluation approaches aim at quantitatively (via quality-changes) and qualitatively (via changes in the PLQ-maps) investigating the effectiveness of the proposed PLQA approach. Lastly, the PLQ-maps are investigated on ICAO-incompliant faces.

[2]https://github.com/davidsandberg/facenet

[3]https://github.com/deepinsight/insightface

[4]We recommend to adjust γ based on an ICAO compliant [20] face image such that the center of the face shows the high quality color (green) while the background shows a uniform color for low quality (red).



(a) ArcFace (b) FaceNet

Fig. 3. **Quality values before and after inpainting** - The values are sorted and plotted with their standard deviation (STD) of calculating the FIQ score 10 times. For ArcFace (FaceNet), the quality scores decreased in 15% (5%) of the cases while the inpainting improved the image quality in 65% (69%) of the cases. In the remaining cases, the quality change was within the STD. The decreases might origin from error-prone inpaintings. In general, inpainting low-quality areas improves the FIQ demonstrating that the PLQ values helped the human observer to interpret and determine the quality-decreasing factor.

V. Results and Discussions

A. Analysing PLQ-Change by Enhancing FIQ

Using the Pre- and Inpainted dataset presented in Section IV-A, we examine the pixel-level quality explanation maps of images with occlusions and other quality degradations as well as images where these impairments have been corrected with inpainting. Comparing the images before and after inpainting allows us to make a statement about how well the proposed solution works for the human observer. Since the observer was asked to mask the quality-decreasing factor, the FIQ of the inpainted image should be higher than the original one if the PLQ-map was well-interpretable.

1) Quantitative Analysis: Figure 3 shows the FIQ values before and after the inpainting for two face recognition models. For ArcFace, only 15% of the quality scores decreased while 65% of the scores increased with inpainting low pixel-quality areas. For FaceNet, only 5% of the inpaintings decreased the FIQ and in 69%, the FIQ increased. In the remaining cases, the quality change was negligible since it was within the standard deviation of the FIQ scores. Moreover, many cases of decreased quality scores origin from bad inpaintings showing unreasonable artefacts (see Figure 4(d)). In general, inpainting areas that our method identified as low-quality improves the FIQ demonstrating that the PLQ-maps helped the human observer to successfully determine the quality-decreasing factor. Especially for images with low FIQ, inpaiting low pixel-quality areas lead to strong quality enhancements as shown in Figure 3.

2) Qualitative Analysis: In Figure 4, the PLQ explanation maps for two face recognition models are shown before and after the inpainting. Figure 4(a) shows a recovered cheek area of a face. For both models, the PLQ-map of the original image shows low pixel-qualities in the area of the covered cheek. In the inpainted image, the cheek is recovered and the area is determined as high-quality pixels. In Figure 4(b), a large occlusion covering the lower part of the face is shown. While for the PLQ-map on ArcFace this area is clearly detected, for FaceNet this is only shown as medium quality. After inpainting, this area is recognized as high-quality by both models. Figure 4(c) shows the effect of glasses on the PLQ-maps. For both models, the frame of the glasses is recognized as low quality and removing the glasses lead to high pixel-qualities. Figure 4(d) shows the case of a faulty inpainting. Before the inpainting, the glasses and reflections are shown as low pixel-qualities. After, the method failed to mark the missing eyes as low quality. In Figure 4(e) a small occlusion is shown. For ArcFace, this occlusion is represented more sharply than for FaceNet. However, removing this occlusion leads to high-pixel qualities for both models. Moreover, the hat is sharply estimated as low-pixel quality for both models. Lastly, Figure 4(f) demonstrates the case of multiple occlusions (headgear and beard). Both occlusions are marked as low-quality pixels and after the inpainting, the qualities are increased. These examples demonstrate that the proposed solutions lead to reasonable pixel-level quality estimates.

B. Analysing PLQ-Change by Decreasing FIQ

Using the Random Mask dataset described in Section IV-A, we examine the pixel-level quality explanations by degrading the high-quality images using randomly placed masks. If the PLQ-maps represent the mask areas as low-quality, we can conclude that our solutions can successfully detect such disturbances.

1) Quantitative Analysis: Figure 5 shows the effect of the random masking process on the image- and pixel-level qualities for two face recognition models and five mask sizes. In Figures 5(a) and 5(b), the distribution of image quality changes affected by the image degradation is shown. The image quality change

$$\Delta_{\hat{Q}} = \hat{Q}_{I_{org}} - \hat{Q}_{I_{mask}} \tag{7}$$

represents the difference between the FIQ of an unmodified image I_{org} and a masked image I_{mask}. A positive $\Delta_{\hat{Q}}$ indicates that the FIQ is successfully degraded in presence of the mask. For the majority of the Random Mask dataset images, such positive values of $\Delta_{\hat{Q}} > 0$ are observed for ArcFace. For FaceNet, low mask sizes ($s = 10, 20$) do not particularly affect the FIQ. Only for medium or larger mask sizes ($s = 30, 40, 50$) the FIQ is successfully degraded.

In Figures 5(c) and 5(d), the distribution of the mean pixel-quality change in the masked area is shown. The mean pixel-quality change

$$\Delta_p = \frac{1}{|\mathcal{P}|} \sum_{i,j \in \mathcal{P}} p_{i,j}^{org} - p_{i,j}^{mask} \tag{8}$$

measures the average difference in the pixel-level qualities $p_{i,j}$ of the unaltered and the masked images in the masked area $\mathcal{P}$. Similar to the image quality change $\Delta_{\hat{Q}}$, a positive Δ_p indicates that the masks lead to degraded pixel qualities. For ArcFace, a large portion of the distributions has positive values for all mask sizes. For FaceNet, this behaviour is only observed for large mask sizes ($s \geq 30$) since the utilized model tends to be robust against smaller disturbances (see Figures 5(b) and 5(d)). In general, the proposed methodology catches the added disturbances with both models and assigns them with significantly lower pixel-level qualities, demonstrating that the produced pixel-level qualities are meaningful.

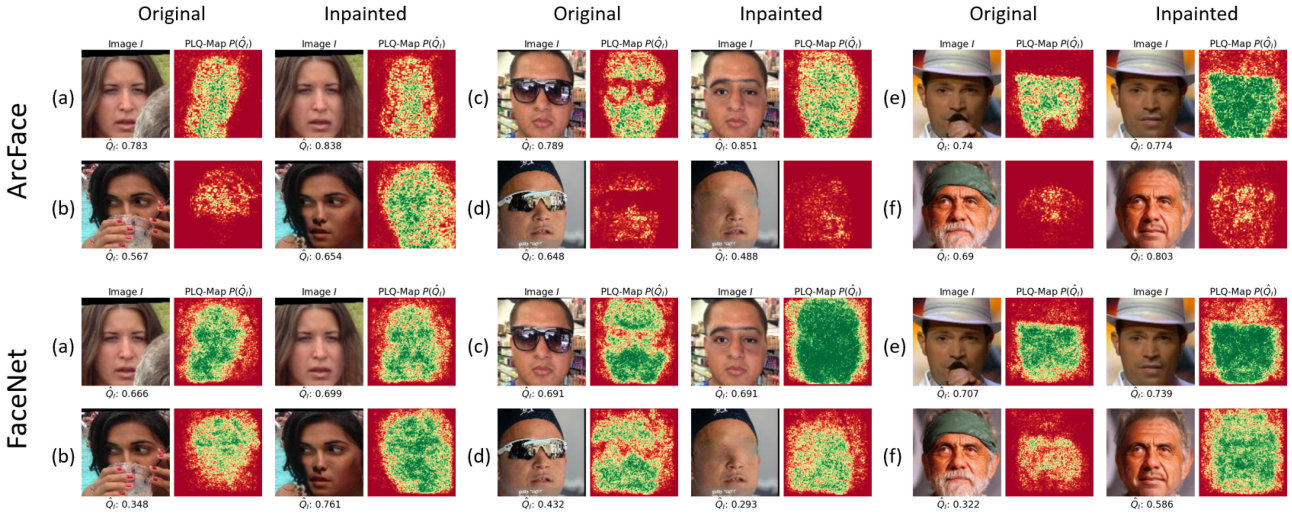
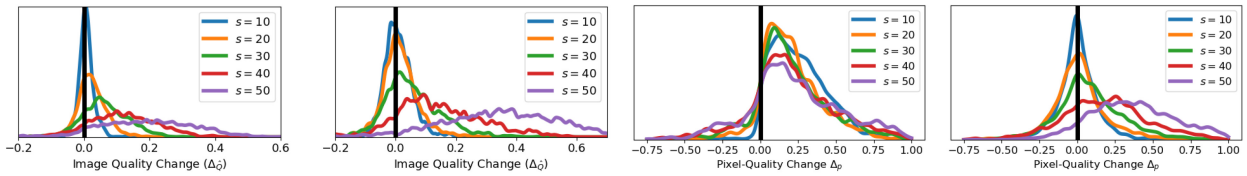


Fig. 4. **PLQ explanation maps before and after inpainting** - Images before and after the inpainting process are shown with their corresponding PLQ-maps and FIQ values. The images show the effect of small and large occlusions, glasses, headgears, and beards on the PLQ-maps for two FR models. In general, these are identified as areas of low pixel-quality and inpainting these areas strongly increases the pixel-qualities of these areas as well as the FIQ. This demonstrates that our solution leads to reasonable pixel-level quality estimates and thus can give interpretable recommendations on the causes of low quality estimates.



(a) Image Quality Change ArcFace (b) Image Quality Change FaceNet (c) Pixel Quality Change ArcFace (d) Pixel Quality Change FaceNet

Fig. 5. **Quality changes through random masks** - High-quality images are degraded by placing random masks of size $s \times s$ pixels on the images. The effect of this is analysed in terms of FIQ change $\Delta_{\hat{Q}}$ of the image and in terms of mean pixel quality change Δ_p in the masked area. The distributions of the image quality changes for both models are shown in (a, b) and (c, d) present the distribution for the pixel quality changes. Positive quality changes (values right of the black line) indicate that the disturbances degrade the qualities. Since the majority of the changes are positive, our solution is able to detect these disturbances and assigns them with low-qualities.

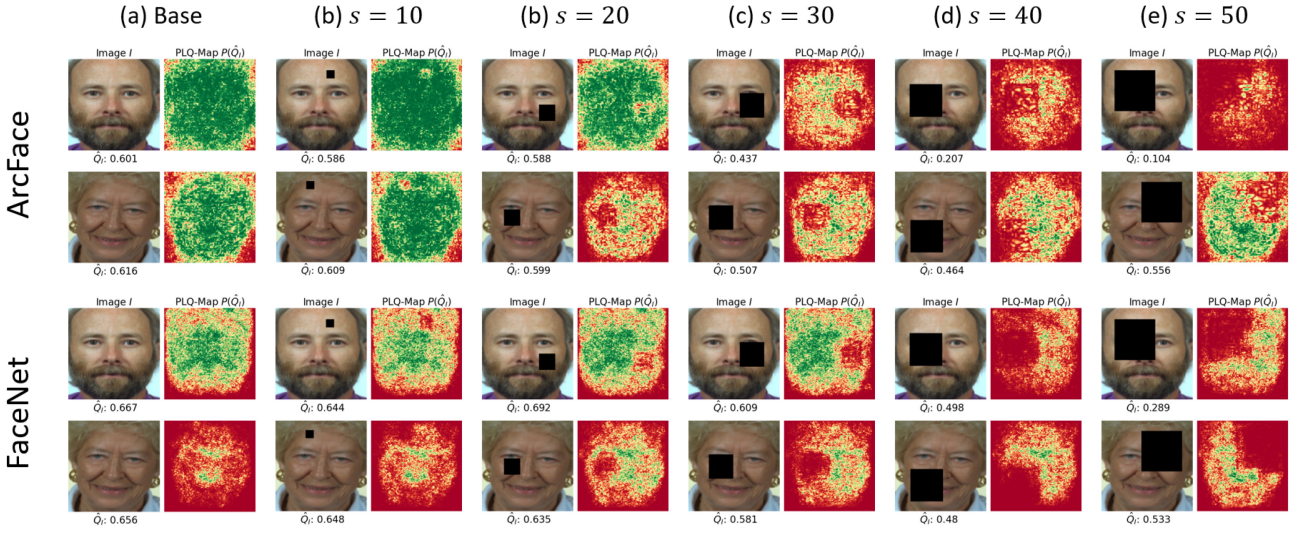


Fig. 6. **PLQ-explanation maps for random masks** - For two random identities, their masked and unmasked images are shown with their corresponding PLQ-maps. In general, the effect of the mask on the PLQ-map is clearly visible demonstrating the effectiveness of the proposed approach to detect disturbances.

2) Qualitative Analysis: In Figure 6, the PLQ explanation maps for two random identities are shown over several mask sizes and locations for both FR models. Moreover, the face images and their PLQ-maps are shown without masks. For all mask sizes, the masked area is assigned with significantly smaller pixel-level quality values than the surrounding pixels.

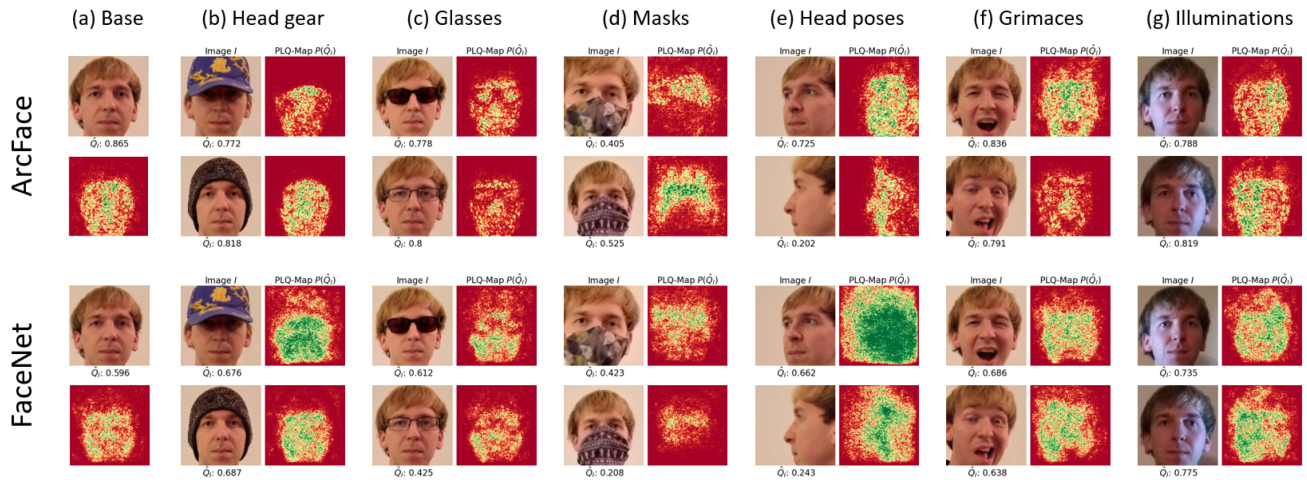


Fig. 7. **PLQ-explanation maps for ICAO incompliant images** - One ICAO-compliant image and twelve images with incompliances are shown with their corresponding PLQ-maps. Occlusions (b, c, d), distorted areas of the face (f), and reflections result in low-pixel qualities.

Consequently, the effect of the masks on the PLQ-map is clearly visible demonstrating that the proposed PLQ assessment approach can detect such disturbances. Besides, it can be seen that the FIQ depends not only on the size but also on the position of the mask.

C. PLQ-Maps on ICAO-Incompliant Images

Lastly, we analyse the proposed approach by investigating the effect of ICAO incompliances on the PLQ-maps. In Figure 7, the ICAO-compliant and several face images with different violations of these specifications are shown with their corresponding PLQ-maps for both FR models. The shown violations include wearing headgear, glasses, and masks, non-frontal head poses, non-neutral expressions, and irregular illuminations. For the ICAO-compliant reference image (a), the area of the face is clearly visible and the PLQ-maps show high pixel-qualities in this area. The same goes for wearing headgears (b) and masks (d) except that the occluded part of the face is assigned with low pixel-quality values. For glasses (c), low-pixel qualities are assigned at the frame of the glasses while the darkened eyes are assigned with higher quality values. Also for the non-frontal head poses (e), the distinction between background and face is clearly visible in the PLQ-maps. Considering non-neutral expressions (f), areas that are distorted compared to neutral expressions are marked as low pixel-quality. Lastly, an interesting effect is observed for non-uniform illuminations. Not the illumined side of the face is assigned with higher quality, instead, the reflections might result in lower pixel-level qualities and the side away from the light is assigned with higher-quality values. Generally, each of the ICAO incompliances can be interpreted with the PLQ maps.

D. Summary

The investigations quantitatively and qualitatively demonstrated the effectiveness of the proposed PLQ assessment approach from two opposite directions and by comparing it with ICAO incompliances. First, in images with low FIQ, low-pixel quality regions are detected and it was shown that inpainting these low-pixel quality regions lead to an increased FIQ. Consequently, the assessment of low pixel-quality areas with the proposed method was correct. Second, images with high FIQ were degraded by placing random masks on the images. The PLQ-maps reliably assigned low-pixel qualities to the areas of disturbances demonstrating the effectiveness of the PLQ assessment. Lastly, the usefulness of the method was proven by showing that the method can detect ICAO incompliances.

Generally, it was shown that the FIQ of an image not only depends on the size of the disturbances (masks) but also on their location on the face. The same applies to PLQ values. The pixel-qualities do not contribute equally and independently to the overall FIQ. Instead, the composition of pixel-qualities might have a stronger impact on the FIQ and thus, on the FR performance. Consequently, a higher FIQ does not necessarily imply that the image must contain more high-quality pixels.

Please note that the proposed analysis focused on the effect of the pixel-level quality on the face image quality. For a comprehensive analysis of the effect of the pixel-level qualities on the recognition performance, we refer to [19]. There, also two evaluation approaches are presented that allow direct comparisons with new pixel-level quality assessment approaches.

Lastly, the PLQ-maps are dependent on the utilized FR models and how it is trained. In our case, the face recognition performance, and thus the FIQ assessment performance, of ArcFace is higher than for FaceNet [39]. Consequently, the quality values and the PLQ-maps are more stable and precise on the ArcFace model. Future works may analyse how well model biases (e.g., for demographics) might be reflected in the PLQ-maps due to the use of model-specific FIQ estimations [38].

VI. Limitations and Ethical Considerations

When applying the proposed methodology we propose to use Gradient Clipping [47] for the backpropagation of

quality-based gradients. This aims to avoid exploding gradients and thus, unreasonable PLQ-maps. Please note that, similar to FR systems, the FIQ on image and pixel-level can be manipulated through adversarial noise. Since FR systems are vulnerable to adversarial noise and model-specific face image quality assessment methods [9], [14], such as SER-FIQ [39] on image-level or the proposed PL-FIQ on pixel-level, aim to estimate the utility of an input for recognition with the deployed system, the quality assessment methods vulnerable to adversarial noise as well. Moreover, we want to emphasize that, depending on the application, inpainting should not carelessly be used to improve FIQ of face images since it might add artefacts leading to wrong matching decisions [25].

VII. Conclusion

The high performance of current FR systems is driven by the quality of its samples. To ensure a high sample quality, for instance, in an automated border control scenario, the FIQ of a captured face is determined. Consequently, a captured face might be rejected during enrolment without a hint of the quality-decreasing factor. In this work, we proposed a methodology to compute pixel-level quality explanation maps to determine which regions of the face have a high and low utility for recognition. Therefore, the proposed approach provides feedback on the utility of a face image that is understandable for humans. Given an arbitrary FR network, we propose a training-free approach that determines the pixel-level quality maps for a face image in three steps. In the first step, a model-specific quality estimate for the image is calculated, modified, and used, in the second step, to construct a quality regression model for the input image. In the third step, quality-based gradients are back-propagated through the model and converted into pixel-level quality maps. The experiments qualitatively and quantitatively demonstrated the effectiveness of the proposed approach in estimating pixel-level qualities. This was shown on real and artificial disturbances and by comparing to ICAO-incompliant images. Moreover, the experiments allowed us to gain more insights into the functionality of FR systems. For instance, it was shown that well-illuminated areas of the face get assigned with significantly lower pixel-qualities than the shaded area of the face. Consequently, the shaded areas provide more important information for the FR models. To summarize, the proposed approach can be applied to arbitrary FR networks, does not require training, and provides a pixel-level utility description of the input face that can be used to (a) deepen the understanding of how FR systems work, (b) enhance the performance of these systems, and (c) to provide understandable feedback of why an image is accepted or rejected during enrolment due to quality concerns.

Acknowledgment

Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office. This work was carried out during the tenure of an ERCIM ‘Alain Bensoussan’ Fellowship Programme.

References

[1] A. Abaza, M. A. Harrison, and T. Bourlai, “Quality metrics for practical face recognition,” in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 3103–3107.

[2] A. Abaza, M. A. Harrison, T. Bourlai, and A. Ross, “Design and evaluation of photometric image quality measures for effective face recognition,” *IET Biometrics*, vol. 3, no. 4, pp. 314–324, 2014.

[3] G. Aggarwal, S. Biswas, P. J. Flynn, and K. W. Bowyer, “Predicting performance of face recognition systems: An image characterization approach,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Colorado Springs, CO, USA, Jun. 2011, pp. 52–59.

[4] D. Baehrens, T. Schroeter, S. Harmeling, M. Kawanabe, K. Hansen, and K.-R. Müller, “How to explain individual classification decisions,” *J. Mach. Learn. Res.*, vol. 11, pp. 1803–1831, Aug. 2010.

[5] L. Best-Rowden and A. K. Jain, “Learning face image quality from human assessments,” *IEEE Trans. Inf. Forensics Security*, vol. 13, pp. 3064–3077, 2018.

[6] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “VGGFace2: A dataset for recognising faces across pose and age,” in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Xi’an, China, May 2018, pp. 67–74.

[7] J. Chen, Y. Deng, G. Bai, and G. Su, “Face image quality assessment based on learning to rank,” *IEEE Signal Process. Lett.*, vol. 22, no. 1, pp. 90–94, Jan. 2015.

[8] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive angular margin loss for deep face recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4685–4694.

[9] Y. Dong et al., “Efficient decision-based black-box adversarial attacks on face recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 7714–7722.

[10] A. Dutta, R. N. J. Veldhuis, and L. J. Spreeuwers, “A Bayesian model for predicting face recognition performance using image quality,” in *Proc. IEEE Int. Joint Conf. Biometrics Clearwater (IJCB)*, Clearwater, FL, USA, Sep./Oct. 2014, pp. 1–8.

[11] E. Eidinger, R. Enbar, and T. Hassner, “Age and gender estimation of unfiltered faces,” *IEEE Trans. Inf. Forensics Security*, vol. 9, pp. 2170–2179, 2014.

[12] M. Ferrara, A. Franco, D. Maio, and D. Maltoni, “Face image conformance to ISO/ICAO standards in machine readable travel documents,” *IEEE Trans. Inf. Forensics Security*, vol. 7, pp. 1204–1213, 2012.

[13] X. Gao, S. Z. Li, R. Liu, and P. Zhang, “Standardization of face image sample quality,” in *Proc. Int. Conf. Adv. Biometrics (ICB)*, Seoul, South Korea, Aug. 2007, pp. 242–251. [Online]. Available: https://doi.org/10.1007/978-3-540-74549-5_26

[14] G. Goswami, N. K. Ratha, A. Agarwal, R. Singh, and M. Vatsa, “Unravelling robustness of deep learning based face recognition against adversarial attacks,” in *Proc. 32nd AAAI Conf. Artif. Intell. (AAAI) 30th Innov. Appl. Artif. Intell. (IAAI) 8th AAAI Symp. Educ. Adv. Artif. Intell. (EAAI)*, New Orleans, LA, USA, Feb. 2018, pp. 6829–6836.

[15] J. Guo, J. Deng, N. Xue, and S. Zafeiriou, “Stacked dense u-nets with dual transformers for robust face alignment,” in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Sep. 2018, p. 44.

[16] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, “MS-celeb-1M: A dataset and benchmark for large-scale face recognition,” in *Proc. 14th Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 87–102.

[17] J. Hernandez-Ortega, J. Galbally, J. Fiérrez, R. Haraksim, and L. Beslay, “FaceQnet: Quality assessment for face recognition based on deep learning,” in *Proc. IEEE Int. Conf. Biometrics (ICB)*, Crete, Greece, Jun. 2019, pp. 1–8.

[18] R.-L. V. Hsu, J. Shah, and B. Martin, “Quality assessment of facial images,” in *Proc. Biometrics Symp. Spec. Session Res. Biometric Consortium Conf.*, Sep. 2006, pp. 1–6.

[19] M. Huber, P. Terhöst, F. Kirchbuchner, N. Damer, and A. Kuijper, “On evaluating pixel-level face image quality assessment,” in *Proc. 30th Eur. Signal Process. Conf. (EUSIPCO)*, Belgrade, Serbia, Aug./Sep. 2022, pp. 1052–1056.

[20] *Machine Readable Travel Documents*, Standard Doc 9303, 2015.

[21] *Information Technology—Biometric Data Interchange Formats—Part 5: Face Image Data*, Standard ISO/IEC 19794-5:2011, Nov. 2011.

[22] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 1867–1874.

[23] H.-I. Kim, S. H. Lee, and M. R. Yong, “Face image assessment learned with objective and relative face image qualities for improved face recognition,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 4027–4031.

[24] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proc. 15th Eur. Conf. Comput. Vision (ECCV)*, Munich, Germany, Sep. 2018, pp. 89–105. [Online]. Available: https://doi.org/10.1007/978-3-030-01252-6_6

[25] J. Mathai, I. Masi, and W. AbdAlmageed, "Does generative face completion help face recognition?" in *Proc. Int. Conf. Biometrics (ICB)*, Crete, Greece, Jun. 2019, pp. 1–8.

[26] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "MagFace: A universal representation for face recognition and quality assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14225–14234.

[27] K. R. Mopuri, U. Garg, and R. V. Babu, "CNN fixations: An unraveling approach to visualize the discriminative image regions," *IEEE Trans. Image Process.*, vol. 28, pp. 2116–2125, 2019.

[28] F.-Z. Ou et al., "SDD-FIQA: Unsupervised face image quality assessment with similarity distribution distance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7670–7679.

[29] P. J. Phillips et al., "On the existence of face quality measures," in *Proc. IEEE 6th Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, Sep. 2013, pp. 1–8.

[30] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[31] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fiérrez, and C. Busch, "Face image quality assessment: A literature survey," 2020, *arXiv:2009.01103*.

[32] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 815–823.

[33] H. Sellahewa and S. A. Jassim, "Image-quality-based adaptive face recognition," *IEEE Trans. Instrum. Meas.*, vol. 59, pp. 805–813, 2010.

[34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, 2020.

[35] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," in *Proc. 2nd Int. Conf. Learn. Represent. (ICLR)*, Banff, AB, Canada, Apr. 2014, pp. 1–8.

[36] D. Smilkov, N. Thorat, B. Kim, F. B. Viégas, and M. Wattenberg, "SmoothGrad: Removing noise by adding noise," 2017, *arXiv:1706.03825*.

[37] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller, "Striving for simplicity: The all convolutional net," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–10.

[38] P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "Face quality estimation and its correlation to demographic and non-demographic bias in face recognition," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Houston, TX, USA, Sep./Oct. 2020, pp. 1–11.

[39] P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 5650–5659.

[40] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, Mar. 2021.

[41] P. Wasnik, K. B. Raja, R. Ramachandra, and C. Busch, "Assessing face image quality for smartphone based face recognition system," in *Proc. 5th Int. Workshop Biometrics Forensics (IWBF)*, Apr. 2017, pp. 1–6.

[42] J. R. Williford, B. B. May, and J. Byrne, "Explainable face recognition," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 248–263.

[43] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *Proc. CVPR Workshops*, Jun. 2011, pp. 74–81.

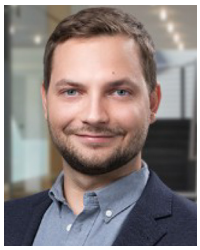
[44] W. Xie, J. Byrne, and A. Zisserman, "Inducing predictive uncertainty estimation for face recognition," 2020, *arXiv:2009.00603*.

[45] B. Yin, L. Tran, H. Li, X. Shen, and X. Liu, "Towards interpretable face recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct./Nov. 2019, pp. 9347–9356.

[46] T. Zee, G. Gali, and I. Nwogu, "Enhancing human face recognition with an interpretable neural network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 514–522.

[47] J. Zhang, T. He, S. Sra, and A. Jadbabaie, "Why gradient clipping accelerates training: A theoretical justification for adaptivity," in *Proc. 8th Int. Conf. Learn. Represent. (ICLR)*, Addis Ababa, Ethiopia, Apr. 2020, pp. 1–21. [Online]. Available: OpenReview.net

Philipp Terhörst received the Ph.D. degree in computer science from the Technical University of Darmstadt in 2021, for his work on "Mitigating Soft-Biometric Driven Bias and Privacy Concerns in Face Recognition Systems" and worked with Fraunhofer IGD from 2017 to 2022. He is a Research Group Leader with Paderborn University working on "Responsible AI for Biometrics." He was also an ERCIM Fellow with the Norwegian University of Science and Technology funded by the European Research Consortium for Informatics and Mathematics. His interest lies in responsible machine learning algorithms in the context of biometrics. This includes the topics of fairness, privacy, explainability, uncertainty, and confidence. He is the author of several publications in conferences and journals, such as CVPR and IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and regularly works as a Reviewer for IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, PR, BTAS, and ICB. For his scientific work, he received several awards, such as from the European Association for Biometrics and the International Joint Conference for Biometrics. He furthermore participated in the "Software Campus" Program and a Management Program of the German Federal Ministry of Education and Research (BMBF).



Marco Huber received the first M.Sc. degree in computer science and the second M.Sc. degree in Internet- and Web-based systems from the Technical University of Darmstadt in 2021. He is a Research Fellow with the Department for Smart Living & Biometrics Technologies, Fraunhofer IGD. He is currently working on secure identity management in biometric systems.

Naser Damer (Member, IEEE) received the Ph.D. degree in computer science from TU Darmstadt in 2018. He is a Senior Researcher with Fraunhofer IGD. He is a Research Area Co-Coordinator and a Principal Investigator with the National Research Center for Applied Cybersecurity ATHENE, Germany. He lectures on Human and Identity-Centric Machine Learning with TU Darmstadt, Germany. His main research interests lie in the fields of biometrics and human-centric machine learning. He serves as an Associate Editor for *Pattern Recognition* (Elsevier) and the *Visual Computer* (Springer). He represents the German Institute for Standardization (DIN) in the ISO/IEC SC37 International Biometrics Standardization Committee. He is a member of the organizing teams of several conferences, workshops, and special sessions, including being the program co-chair of BIOSIG and a member of the IEEE Biometrics Council serving on its Technical Activities Committee.

Florian Kirchbuchner (Member, IEEE) received the Master of Science degree in computer science from the Technical University of Darmstadt in 2014, where he is currently pursuing the Ph.D. degree on the topic "Electric Field Sensing for Smart Support Systems: Applications and Implications." He has been working with the Fraunhofer IGD since 2014, most recently as the Head of the Department for Smart Living & Biometric Technologies. He is also a Principal Investigator with the National Research Center for Applied Cybersecurity ATHENE. He participated at Software Campus, a Management Program of the Federal Ministry of Education and Research. He is trained as an Information and Telecommunication Systems Technician and served as an IT Expert for German Army from 2001 to 2009.

Kiran Raja (Senior Member, IEEE) received the Ph.D. degree in computer Science from the Norwegian University of Science and Technology, Norway, in 2016, where he is a Faculty Member with the Department of Computer Science. He was/is participating in EU projects SOTAMD, iMARS, and other national projects. His main research interests include statistical pattern recognition, image processing, and machine learning with applications to biometrics, security and privacy protection. He is a member of European Association of Biometrics (EAB) and the Chair of Academic Special Interest Group, EAB. He serves as a reviewer for number of journals and conferences. He is also a member of the editorial board for various journals.

Arjan Kuijper (Member, IEEE) received the M.Sc. degree in applied mathematics from Twente University, The Netherlands, the Ph.D. degree from Utrecht University, The Netherlands, and the habitation degree from TU Graz, Austria. He holds the Chair in "Mathematical and Applied Visual Computing" with TU Darmstadt and is a member of the Management of Fraunhofer IGD, responsible for scientific dissemination. He was an Assistant Research Professor with the IT University of Copenhagen, Denmark, and a Senior Researcher with RICAM, Linz, Austria. He is the author of over 350 peer-reviewed publications. His research interests cover all aspects of mathematics-based methods for computer vision, biometry, graphics, imaging, pattern recognition, interaction, and visualization. He is an Associate Editor for *Computer Vision and Image Understanding*, PR, and TVCJ, the President of the International Association for Pattern Recognition, serves as a reviewer for many journals and conferences, and as well as a program committee member and an organizer of conferences.